

W200 Project 2 Proposal

Team Members: Jared Dec, Brenna Giacchino, Aidan Jackson

GitHub Repository: Project2_Dec_Giacchino_Jackson

https://github.prod.oc.2u.com/UCB-INFO-PYTHON/Project2_Dec_Giacchino_Jackson

Research Question:

How do the demographics of the voting age population in states in the U.S. compare to their elected officials?

Analysis Overview:

The purpose of this project is to determine how the demographics of elected officials may compare to the voting age population which elect them. While there are various levels of elected officials, this analysis will focus on those who hold local positions at the town or city level. Local elections have been noted for both how “scale...can affect political participation”^[2] as well as being “the nation’s foremost venues for the study of political behavior in the context of significant racial and ethnic diversity”.^[1] Considering this, voters in local elections may be motivated by the greater influence their vote would have relative to the total voting population, or also by candidates who are from their community as opposed to being potentially distant.

While the scope of this project will not be to determine any causal relationships between voters and elected officials, it will investigate whether elected officials at the local level are representative of the voter populations they represent. This will be done via a chi-squared hypothesis test, where the demographics of local elected officials across each state are compared to demographics of voters in that state. Although the elected positions are local, analysis will be conducted on a state by state basis to increase the degrees of freedom for each statistical test. The test will proceed under a null hypothesis that there is no difference between the demographics of the statewide voters and their local elected officials, with an alternative hypothesis that there is a difference (corresponding to a two-tailed test). Following the results from each state, the data will be further examined to determine if states which are significant for one variable, such as race, are also correlated with being significant for another variable, such as sex.

Datasets:

The primary [data set](#) is provided by the [Reflective Democracy Campaign](#), a group which investigates and promotes women and people of color in elected offices across the United States. The group collects data around demographics of elected officials from local to national governments, as well as conducts research to determine how systemic barriers may be overcome by these candidates. Analysis will center around a collection of local officials in the 100 most populous cities in the US that were in office in September of 2020. The primary positions occupied by the local officials are city council members and mayors, although other elected positions such as City Attorney or City Clerk are included if applicable. The scope of the elected officials in the data set is restricted to either the city or county level.

A supplemental dataset is required in order to have information on the demographics of the voter population in each state for comparison with the data on the elected officials. This was

obtained from a US Census Bureau [publication](#) on 2018 voter-age demographics. Of note is that this dataset only estimates the population of voting-age US citizens for each state, and not the population of registered/eligible voters or those who were verified as having actually voted. According to one study from the Pew Research Center from 2014, approximately 21.4% of eligible voters were not actually registered to vote.³ This dataset is also two years older than the primary dataset, which will require the assumption that demographics in the voting-age US citizen population of each state has not meaningfully changed in that time.

Relevant Variables:

Primary Dataset

For the occupant of each local elected position, information on the occupant's race, sex, and city/state is included. The "Hispanic or Latino" category is included within the race variable, along with "Multiracial" and "Unknown" categories. This choice for the "Hispanic or Latino" category is in contrast with including a separate variable just for that category, such as "Ethnicity". In addition, the sex variable is restricted to two categories, either female or male. Based on provided documentation, it is unknown how data collection may or may not have accounted for intersex people. These elected official demographics will be aggregated by state for comparison to voter populations.

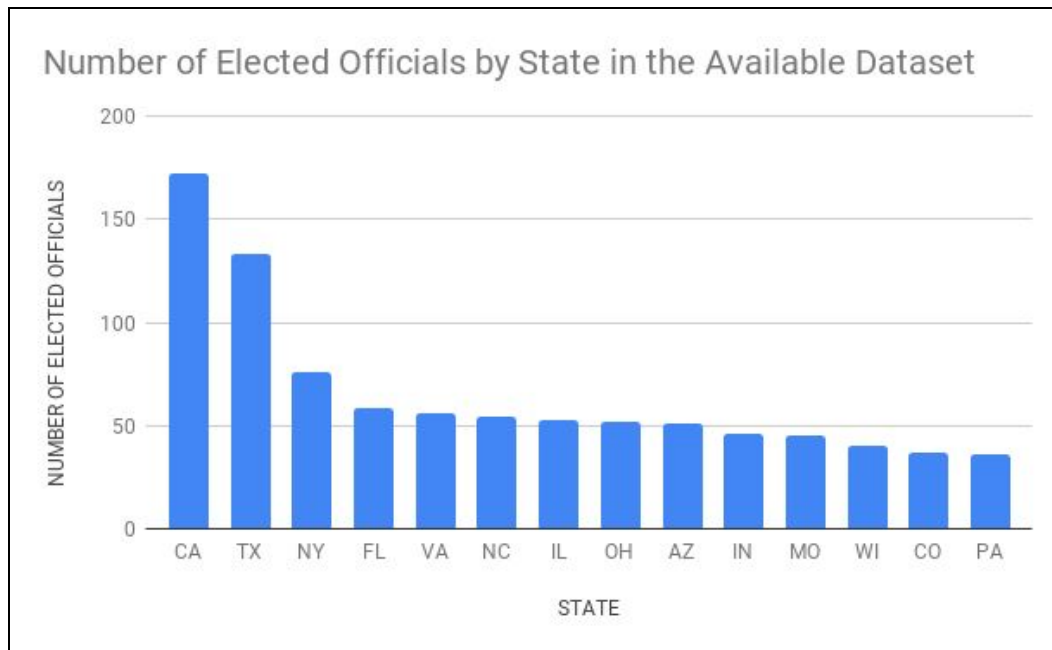
Supplemental Dataset

For each state, estimates of the number of voting age US citizens and an associated margin of error is presented for each of twelve racial categories. This includes categories which are similar as the primary dataset, although they are presented in different formats. For example, several multiracial categories are presented distinctly, such as "Asian and White", "Black or African American and White", or "Remainder of Two or More Race Responses". Similarly to the primary dataset, however, the "Hispanic or Latino" category is also considered mutually exclusive with other race categories.

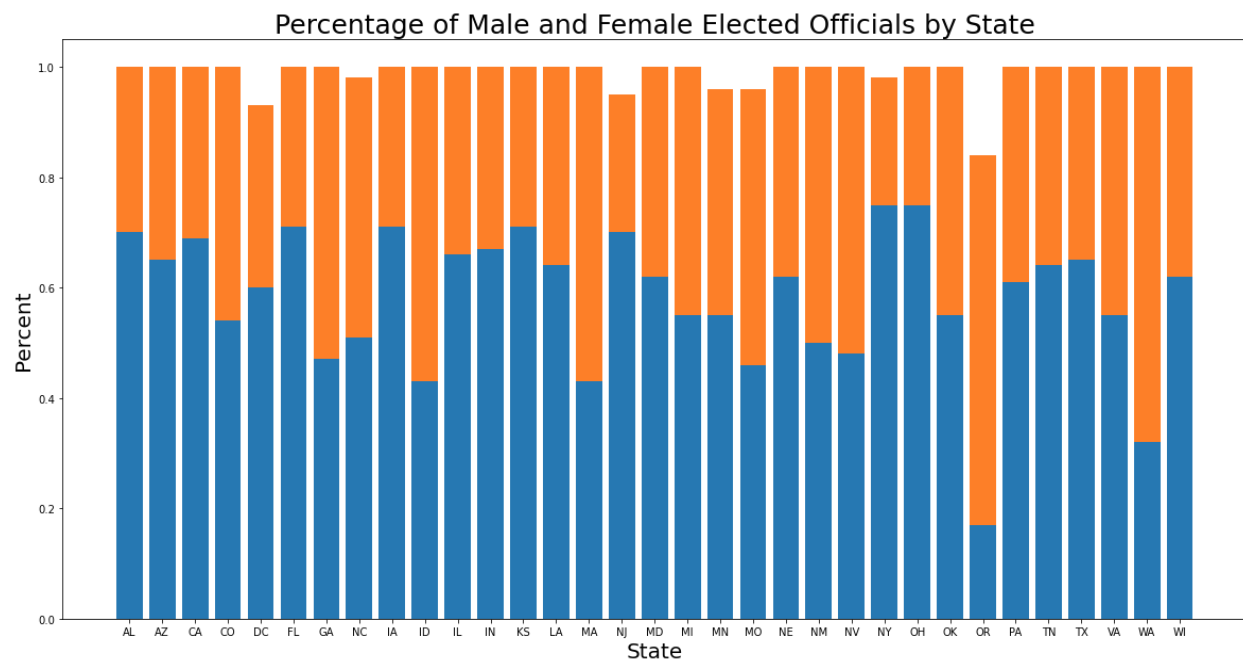
Exploratory Data Analysis:

The plan for implementing data analysis into this project involves creating two sets of data that encompass the percent composition of demographics for the voting age population by race and sex and the percent composition for the same categories for elected officials for the same constituency. From there, a chi-squared test could be used to determine the statistical significance of the difference in distribution between the two datasets. Other than manipulating the source datasets to get them in a position for a chi-squared test to be possible, this will be the sole data analytical tool employed. The chi-squared test will be used for differing levels of constituency in order to map out what particular regions have political representation most in line with the underlying population.

Initial Charts:



This chart displays the top 14 states in the data set with the highest number of reported elected officials. These are the states that we will most likely be using for our chi squared testing.



Note: There should be a legend for this chart that indicates the blue bars represent the males and the orange bars represent the females. Some state percentages don't add up to 1.0 due to the sex of some elected officials being "Unknown".

Works Cited:

[1] Kaufmann, Karen M. & Rodriguez, Antonio. Political Behavior in the Context of Racial Diversity: The Case for Studying Local Politics. University of Maryland.

<https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/3/247/files/2016/07/SYMPOSIUM-Marschall-Shah-Ruhil-1520gbn.pdf>

[2] Lewis, Paul G. Size and Local Democracy: Scale Effects in City Politics. Arizona State University.

<https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/3/247/files/2016/07/SYMPOSIUM-Marschall-Shah-Ruhil-1520gbn.pdf>

[3]

Why Are Millions of Citizens Not Registered to Vote?, The Pew Charitable Trusts, 21 June 2017,

www.pewtrusts.org/en/research-and-analysis/issue-briefs/2017/06/why-are-millions-of-citizens-not-registered-to-vote.